RESPIRATORY PATTERN VARIABILITY ANALYSIS BASED ON NONLINEAR PREDICTION METHODS

L. Domingo[1], P. Caminal[1], B.F. Giraldo[1], S. Benito[2], M. Vallverdú[1], D. Kaplan[3]

[1]Centre de Recerca en Enginyeria Biomèdica (CREB), ESAII, Univ. Politècnica Catalunya, Barcelona, Spain
[2]Dep. Intensive Care Medicine, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.
[3]Dep. Mathematics and Computer Science, Macalester College, St. Paul, Minnesota, USA

Abstract – **The traditional techniques of data analysis are often not sufficient to characterize the complex dynamics of respiration. In this study the respiratory pattern variability at different levels of pressure support ventilation (PSV) has been analyzed using nonlinear prediction methods. These methods use the volume signals generated by the respiratory system in order to construct a model of its dynamics, and then to estimate the deterministic level of the system from the quality of the predictions made with the model. Different methods of prediction evaluation and neighborhood definition have been considered. The incidence of different prediction depths and embedding dimensions have been analyzed. A group of 12 patients on weaning trials from mechanical ventilation have been studied at two different PSV levels. High statistically significant differences have been obtained when comparing the mean prediction error at two different PSV levels ($p<0.002$) with non-parametric analysis of variance test (Wilcoxon's signed rank test). The embedding dimension needed to model the system dynamics with low prediction error has also presented significant differences ($p<0.005$) between the complex dynamics of both PSV levels. Therefore, it may be concluded that the respiratory pattern variability depends on the level of pressure support ventilation.**

Keywords – **Respiratory pattern variability, nonlinear prediction methods, pressure support ventilation.**

I. Introduction

The possible causes of breath-to-breath variability in the pattern of breathing have been discussed recently [1]. The traditional techniques of data analysis in the time and frequency domains are often not sufficient to characterize the complex dynamics of respiration. Various attempts have been reported to apply the concept of nonlinear dynamics to the analysis of complex physiological systems [2-4] and to distinguish between variations that are random and those that are deterministic ("chaotic"). It has been shown that chaotic measurements like correlation dimension present an irregular behavior of the respiratory system in the wakefulness stages and a less complex dynamical structure during sleep [5]. Other studies have demonstrated that the dynamics of infant breathing during quiet sleep can best be described as a chaotic system [6]. In our study the respiratory pattern complexity in different levels of pressure support ventilation has been analyzed using nonlinear prediction methods.

II. Material and Methods

A. Patients and Data Acquisition

A group of 12 patients on weaning trials from mechanical ventilation has been studied in the Department of Intensive Care Medicine at Santa Creu i Sant Pau Hospital. Patients were submitted under two different levels of pressure support ventilation (PSV). The respiratory volume signals were obtained by means of a respiratory inductive plethysmograph. Respiratory volume at each PSV level was recorded during 30 minutes with a sampling frequency of 250 Hz and resampled at 10 Hz for this study.

B. Nonlinear Prediction

One method to decide whether an underlying deterministic system is present is the following: To use the time series generated by the system in order to construct a model of the dynamics, and then to see whether the predictions made from this model are accurate. If the predictions are perfect, then the system is completely deterministic. If the predictions are good, but not perfect, then the system has a deterministic component. If the predictions are bad, then the system is not deterministic at all [7].

There are different ways to construct dynamical models from data. Since all of the state variables of the systems are not directly measured, the embedding technique to represent all of the measured data's state variables has been used. By embedding the scalar time series D_t, the following vector sequence is created:

$$\boldsymbol{D}_t = (D_t, D_{t-\tau}, \ldots D_{t-(m-1)\tau})$$

where m is the embedding dimension and τ is the embedding lag. Each $\boldsymbol{D_t}$ is a point in the m – dimensional embedding space, and the embedded time series can be regarded as a sequence of points, one point at each time t. Each point represents the state of the system at that time.

A deterministic data set sampled at discrete times can be described by a discrete-time map

$$\boldsymbol{D}_{t+1} = \boldsymbol{F}(\boldsymbol{D}_t)$$

Report Documentation Page

Report Date	Report Type	Dates Covered (from... to)
25 Oct 2001	N/A	-

Title and Subtitle Respiratory Pattern Variability Analysis Based on Nonlinear Prediction Methods	**Contract Number**
	Grant Number
	Program Element Number
Author(s)	**Project Number**
	Task Number
	Work Unit Number
Performing Organization Name(s) and Address(es) Centre de Recerca en Enginyeria Biomedica (CREB), ESAIL, Univ. Politecnica Catalunya, Barcelona, Spain	**Performing Organization Report Number**
Sponsoring/Monitoring Agency Name(s) and Address(es) US Army Research, Development & Standardization Group (UK) PSC 802 Box 15 FPO AE 09499-1500	**Sponsor/Monitor's Acronym(s)**
	Sponsor/Monitor's Report Number(s)

Distribution/Availability Statement
Approved for public release, distribution unlimited

Supplementary Notes
Papers from 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, October 25-28, 2001, held in Istanbul, Turkey. See also ADM001351 for entire conference on cd-rom., The original document contains color images.

Abstract

Subject Terms

Report Classification unclassified	**Classification of this page** unclassified
Classification of Abstract unclassified	**Limitation of Abstract** UU

Number of Pages
4

which is, however, immediately applicable only if the mapping $\boldsymbol{F}$ is known. With $\boldsymbol{F}$ unknown some assumptions about its properties have to be made. With the minimal assumption that the mapping $\boldsymbol{F}$ is continuous the following prediction scheme can be constructed [8]. In order to predict the future state $\boldsymbol{D}_{t+1}$ given the present one $\boldsymbol{D}_t$, the state closest to $\boldsymbol{D}_t$ with respect to some norm is searched. Let's say that this closest point has time index a. The definition of determinism is that future events are set causally by the past events. $\boldsymbol{D}_t$ describes the past events to D_{t+1}. Similarly $\boldsymbol{D}_a$ describes the past events to the measurement D_{a+1}. If $\boldsymbol{D}_t$ is close to $\boldsymbol{D}_a$, and if the system is deterministic, then it is expected that D_{a+1} will also be close to D_{t+1}. In order to predict a time h ahead of t D_{t+h} the vector $\boldsymbol{D}_a$ closest to $\boldsymbol{D}_t$ has to be found and D_{a+h} will be used as a predictor and it will be called P_{t+h}.

Every measurement of a continuous quantity is only valid up to some finite resolution and this fact has to be taken into account. The finite resolution implies that looking for the single closest state is no longer the best can be done since interpoint distances are contaminated with an uncertainty. All points within a close region in phase space have to be considered to be equally good predictions a priori. Then the proposed prediction algorithm to be used forms a neighborhood $U(\boldsymbol{D}_t)$ around the point $\boldsymbol{D}_t$. For all points $\boldsymbol{D}_{a_i} \in U(\boldsymbol{D}_t)$, that is, all points close to $\boldsymbol{D}_t$ look up the individual predictions D_{a_i+h}. The finally accepted prediction is then the average of all these individual predictions:

$$P_{t+h} = \frac{1}{\left|U(D_t)\right|} \sum_{D_{a_i} \in U(D_t)} D_{a_i+h}$$

where $\left|U(D_t)\right|$ denotes the number of elements of the neighborhood $U(\boldsymbol{D}_t)$. Two ways have been considered in order to define the neighborhood: i) the neighbors inside an hypersphere of radius ε around the point $\boldsymbol{D}_t$; ii) the k neighbors closest to the point $\boldsymbol{D}_t$.

Given a method for making a prediction P_{t+h}, an actual measurement of D_{t+h} is needed in order to decide if the prediction is good or bad. The difference between P_{t+h} and D_{t+h} is the prediction error, which informs about the quality of the prediction. As a single prediction might be good or bad just by chance, in order to give a more meaningful indication of the determinism in the data an average of many prediction errors should be taken. Two different ways have been considered in order to define this indication of determinism: i) Nonlinear cross-prediction; ii) Leave-one-out cross validation.

In the nonlinear cross-prediction approach the time series is broken into M segments. For each of the M segments, one at a time, the model is fit and then residuals are calculated on each of the other segments. The residuals are summarized by one number, the mean absolute value. The result is a M-by-M matrix of cross-predictabilities. In this study the respiratory volume data set at each PSV level that contains 18000 samples has been divided in $M=3$ segments of 6000 samples. In this case the 3-by-3 matrix has 6 entries (the diagonal elements that correspond to self prediction are not computed) and their mean value is computed in each patient for each PSV level.

In the leave-one-out cross validation the time series of length N is modeled N different times: for each model, a single data point is left out when fitting the model and the residual for the model is computed only for the left-out data point. The result is a set of residuals one for each point, that provide an estimate of the prediction error of a model. In this study the respiratory volume data set at each PSV level has been divided in $N=9$ subsets of 2000 samples. In this way the mean prediction error related to each patient for each PSV level corresponds to the mean value of the prediction errors in the nine subsets.

A preprocessing step has been applied to each respiratory volume data set in order to improve the analysis of the results. The respiratory volume signals have been subtracted by their mean value and divided by their variances. The embedding technique has been applied using the embedding lag τ corresponding to the first zero of the autocorrelation function [9]. Fig. 1a and 1b show the actual measurements and predictions for the respiratory volume of a patient under low and high PSV, respectively. The different quality of the prediction is shown comparing both PSV levels.

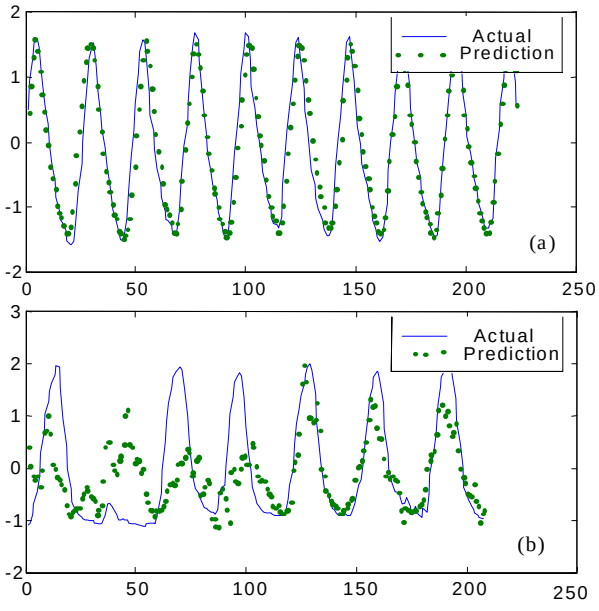

Fig. 1. Respiratory volume (actual measurement and prediction) of a patient with a low (a) and high (b) pressure support ventilation level.

A non-parametric analysis of variance test, Wilcoxon's signed rank test, has been used to analyze statistically the differences between the two PSV levels. This method is

used for testing the difference between two populations using matched samples.

III. Results

The first step was to define the methodology to be applied: nonlinear cross-prediction or leave-one-out cross validation (auto-prediction), the best kind of neighborhood and the best prediction horizon *h*. In this previous study three patients were analyzed, and an embedding dimension $m = 2$ was considered. When deciding between cross-prediction and auto-prediction two kinds of neighborhoods were analyzed: the neighbors inside an hypersphere of radius $\varepsilon = 0.2$ and the $k = 20$ closest neighbors. Tables I and II present as an example the values obtained in patient CRR using the neighbors inside an hypersphere and the *k* closest neighbors, respectively. In the three analyzed patients the statistical significance (p-value) obtained when comparing low and high PSV levels were found not dependent of the auto-prediction or cross-prediction methodologies. Then, as the auto-prediction represents a lower time consuming, this methodology has been selected for the next steps.

TABLE I

MEAN ± STANDARD DEVIATION FOR THE MEAN PREDICTION ERROR OF THE PATIENT CRR WITH $M = 2$, $\varepsilon = 0.2$, WHEN CONSIDERING LEAVE-ONE-OUT CROSS VALIDATION AND NONLINEAR CROSS-PREDICTION. STATISTICAL SIGNIFICANCE: P-VALUE.

	Low PSV	*High PSV*	*p-value*
Leave-one-out validation	0.41 ± 0.05	0.81 ± 0.03	$p < 0.01$
Nonlinear cross-prediction	0.45 ± 0.04	0.90 ± 0.06	$p < 0.01$

TABLE II

MEAN ± STANDARD DEVIATION FOR THE MEAN PREDICTION ERROR OF THE PATIENT CRR WITH $M = 2$, $K = 20$, WHEN CONSIDERING LEAVE-ONE-OUT CROSS VALIDATION AND NONLINEAR CROSS-PREDICTION.

	Low PSV	*High PSV*	*p-value*
Leave-one-out validation	0.38 ± 0.05	0.79 ± 0.02	$p < 0.01$
Nonlinear cross-prediction	0.43 ± 0.04	0.88 ± 0.05	$p < 0.01$

In order to decide the best kind of neighborhood to discriminate the different determinism of the respiratory volume, in low and high PSV levels, the following neighborhoods were considered: the neighbors inside hyperspheres of radius $\varepsilon = 0.1, 0.2, 0.3$ and the $k = 20$ closest neighbors. Table III presents as an example the values obtained in patient CRR. In the three analyzed patients the statistical significance (p-value) obtained when comparing low and high PSV levels were found not dependent of the different neighborhood methodology. Then, as the radius of the hyperspheres could be dependent of the embedding dimension, the *k* closest neighbors methodology has been selected for the next steps.

TABLE III

MEAN ± STANDARD DEVIATION FOR THE MEAN PREDICTION ERROR OF THE PATIENT CRR WITH $M = 2$ WHEN CONSIDERING DIFFERENT RADIUS ε OF THE HYPERSPHERES AND THE $K=20$ CLOSEST NEIGHBORS.

	Low PSV	*High PSV*	*p-value*
$\varepsilon = 0.1$	0.45 ± 0.06	0.85 ± 0.02	$p < 0.01$
$\varepsilon = 0.2$	0.41 ± 0.05	0.81 ± 0.03	$p < 0.01$
$\varepsilon = 0.3$	0.39 ± 0.06	0.80 ± 0.02	$p < 0.01$
k neighbors	0.38 ± 0.05	0.79 ± 0.02	$p < 0.01$

The next analysis has been done to select the best prediction horizon h. For each patient and for each PSV level the mean respiratory period has been calculated. This mean respiratory period translated to sample units is called h_{Ttot}. Three prediction horizons have been considered: 0.5 h_{Ttot}, h_{Ttot} and 2 h_{Ttot}. Table IV presents as an example the values obtained in patient CRR using the different prediction horizons. In the three analyzed patients the statistical significance (p-value) obtained when comparing low and high PSV levels were found not dependent of the considered h value. A prediction depth related to the mean respiratory period has been selected for the next steps.

TABLE IV

MEAN ± STANDARD DEVIATION FOR THE MEAN PREDICTION ERROR OF THE PATIENT CRR WITH $M = 2$ WHEN CONSIDERING DIFFERENT PREDICTION HORIZONS *H*.

	Low PSV	*High PSV*	*p-value*
$0.5\, h_{Ttot}$	0.39 ± 0.07	0.67 ± 0.09	$p < 0.01$
h_{Ttot}	0.38 ± 0.05	0.79 ± 0.02	$p < 0.01$
$2\, h_{Ttot}$	0.55 ± 0.09	0.81 ± 0.08	$p < 0.01$

The incidence of the embedding dimension *m* on the prediction errors has been analyzed in all the patients for each one of the PSV levels. Fig. 2 shows as an example the relation between the mean prediction error and the embedding dimension for the patient CRR. The behavior depends on the PSV level as can be seen in the figure.

In order to analyze the level of determinism in the respiratory volume signals related to high PSV level in comparison with the low PSV level, Table V shows the mean prediction errors obtained for $m = 2$ when considering

all the patients. The results show a statistically significant difference ($p<0.002$) between both groups.

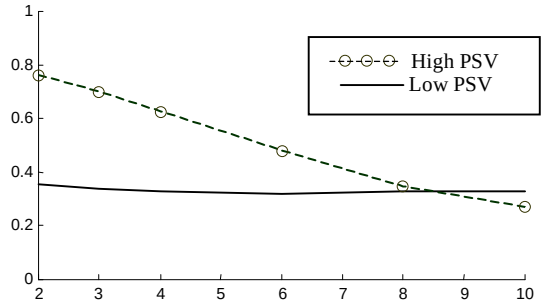

Fig. 2. Prediction errors obtained as a function of the embedding dimension *m* for the patient CRR.

TABLE V

MEAN ± STANDARD DEVIATION OF THE MEAN PREDICTION ERROR FOR ALL THE PATIENTS.

	Low PSV	*High PSV*	*p-value*
Mean prediction error	0.36 ± 0.09	0.59 ± 0.12	$p < 0.002$

Another way to analyze the results is the study of the embedding dimension needed to model the dynamics of the patients with a low prediction error (< 0.40). For example in patient CRR (Figure 2) at high PSV level an embedding dimension $m = 8$ is needed to get a prediction error below 0.4, while a $m = 2$ is enough to get the same prediction error for the low PSV level. Table VI shows the values obtained when analyzing all the patients. The embedding dimension needed to model the dynamics of the patients with a low prediction error show a statistically significant difference ($p<0.005$) between both groups.

TABLE VI

MEAN ± STANDARD DEVIATION FOR THE EMBEDDING DIMENSION NEEDED TO MODEL THE DYNAMICS OF THE PATIENTS WITH A REDUCED PREDICTION ERROR.

	Low PSV	High PSV	p-value
Embedding dimension	2.67 ± 1.30	6.33 ± 2.23	$p < 0.005$

V. DISCUSSION AND CONCLUSIONS

To compare the respiratory pattern variability at two different levels of pressure support ventilation nonlinear prediction methods have been applied. The volume time series have been used to construct a model of the respiratory system dynamics and the accuracy of the predictions made from the model have been analyzed. Two different ways have been considered in order to define the indication of determinism: nonlinear cross-prediction and leave-one-out cross validation. Two kinds of neighborhoods have been analyzed: the neighbors inside an hypersphere of radius ε and the *k* neighbors closed to a point in the phase space. The incidence of different prediction depths has been also considered. The analysis of the prediction error as a function of the embedding dimension has been used to propose a new index to discriminate different respiratory pattern variability levels. High statistically significant differences have been obtained when comparing the mean prediction error at two different PSV levels ($p<0.002$). The embedding dimension needed to model the dynamics of the system with a low prediction error is also a good parameter to discriminate ($p<0.005$) different respiratory patterns. Therefore, it may be concluded that the respiratory pattern variability depends on the level of pressure support ventilation.

ACKNOWLEDGMENT

This work was partially supported by FEDER grant 2FD1997-1197-C02-02 and FIS grant 99/0226.

REFERENCES

[1] E.N. Bruce. Measures of respiratory pattern variability. In: *Bioengineering approaches to pulmonary physiology and medicine*, Plenum Press, 149-160. (1996).

[2] D.R. Rigney, W.C. Ocasio, K.P. Clark, J.Y. Wei and A..L. Goldberger. Deterministic mechanism for chaos and oscillations in heart rate and blood pressure. *Circulation*: 1-659. (1992).

[3] R.G. Turcott, M.C. Teich. Fractal character of the electrocardiogram: distinguishing heart failure and normal patients. *Ann. Biomed. Eng. 24*: 269-293. (1996).

[4] P. Achermann, R. Hartmann, A. Gunzinger, W. Guggenbuhl, A.A. Brobely. All-night sleep EEG and artificial stochastic control signals have similar correlation dimensions. *Electroencephalogr. Clin. Neurophisiol.* 90 : 384-387. (1994).

[5] C. González, P. Caminal, S. Benito. Complexity analysis of the respiratory system. *Proc. 5th Conf. European Society for Engineering and Medicine ESEM99* : 565-566. (1999).

[6] M. Small, K. Judd, M. Lowe, S. Stick. Is breathing in infants chaotic? Dimension estimates for respiratory patterns during quiet sleep. *J. Appl. Phisiol.* 86(1): 359-376. (1999).

[7] D. Kaplan and L. Glass. Characterizing chaos. In *Understanding nonlinear dynamics*. Springer-Verlag: 314-338. (1995).

[8] H. Kantz and T. Schreiber. Determinism and predictability. In *Nonlinear time series analysis*. Cambridge University Press: 42-57. (2000).

[9] H.D. Abarbanel, R. Brown, J.B. Kadtke. Prediction in chaotic nonlinear systems: Methods for time series with broadband Fourier spectra, *Phys. Rev. A*, vol. 41, 1782-1807. (1989).